CHARACTERIZING THE OPTIX NETWORK IN HELICONIUS BUTTERFLY WING COLOR PATTERNING

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Kathleen Elizabeth Rondem

August 2018

ABSTRACT

Heliconius butterflies are an incredibly diverse species, with many displaying bold red, black and yellow coloring. Interestingly, a single gene, *optix*, has been shown to be responsible for the red coloring of these as well as many other butterflies within the Lepidoptera order. Optix is a highly conserved transcription factor with ancestral function in *Drosophila* eye development, so how has this gene evolved its function in butterfly wing pigmentation? Here, I describe work I did to elucidate both the upstream and downstream actors of the Optix network in the *Heliconius* wing pigmentation pathway in an effort to provide insight into how this novel function of *optix* has evolved. First, I explore the role of *optix* regulatory elements in wing color patterning using the CRISPR/Cas9 system to delete candidate *optix* cis-regulatory elements. I then investigate the conservation of the ancestral Optix eye network in the pigmentation pathway by deleting candidate downstream target genes of Optix and describe the resulting wing pattern mutations. This work provides foundational information on the role of this network in *Heliconius* wing color patterning and how it may have evolved.

BIOGRAPHICAL SKETCH

Katie Rondem is a graduate student in the Ecology and Evolutionary Biology Department at Cornell University and works in Dr. Robert Reed's lab investigating the evolution and development of butterfly wing color patterns. She has an A.B.J. degree from the University of Georgia and a B.S. degree from the University of Utah, where she worked in Erik Jorgensen's lab on synaptic vesicle dynamics in the *C.elegans* system. Katie has been active in scientific outreach during her graduate school career, most notably as an animator for Graduate Women in Science's monthly video series, "This Is What A Scientist Looks Like". Following completion of her degree, she will be working at the Huntsman Cancer Institute's Genomics Core Facility in Salt Lake City, UT.

ACKNOWLEDGMENTS

I would like to first acknowledge the people in my lab who helped me immensely in the planning and completing of my thesis work. Thank you to my advisor, Dr. Bob Reed, for providing encouragement and guidance and for always having an open door. Thank you to James Lewis for providing the groundwork needed for me to do the functional experiments I dreamed of doing when coming to grad school and for sharing his work with me. And thank you to my labmates—Anyi Mazo-Vargas, Karin van der Berg, and Linlin Zhang—who provided endless advice and help and never failed in giving me positive support when I needed it the most.

I would also like to acknowledge the wonderful people in the Ecology and Evolutionary Biology Department who made my time at Cornell something special, especially my fellow cohort-mates. Between my lab and my department, I feel very lucky to have spent time with such an incredible group of smart, driven, and kind people.

And finally, I would like to acknowledge my husband, who never faltered in his support throughout this whole process—and without whom I probably would have starved.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 2 FIGURES

CHAPTER 3 FIGURES

LIST OF TABLES

CHAPTER 2 TABLES

CHAPTER 3 TABLES

LIST OF ABBREVIATIONS

3-OHK	3-hydroxykyurenine
ATAC	Assay for Transposase-Accessible Chromatin
ATACseq	Assay for Transposase-Accessible Chromatin with sequencing
Bp	Base pair
Cn	Cinnabar
CRE	Cis-regulatory element
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
HW	Hindwing
INDEL	Insertion/deletion
FW	Forewing
Kb	Kilobase
KO	Knock-out
MFS	Major Facilitator Superfamily
MFS2	Major Facilitator Superfamily 2
mKO	Mosaic knock-out
PCR	Polymerase Chain Reaction
sgRNA	single-guide RNA
SNP	Single nucleotide polymorphism
WT	Wild-type

CHAPTER 1

INTRODUCTION

Heliconius butterflies have been the subject of evolutionary study for many years, and for good reason: they are incredibly diverse, with over 40 species displaying a fantastic array of pattern and color diversity; they participate in complex mimicry rings, lending adaptive significance to their beautiful coloring; and they are relatively easy to cross, providing access to association studies that elucidate important genomic regions underlying their patterning. Through previous mapping and association work, gene expression studies, and, more recently, functional validation work using the CRISPR/Cas9 system, one gene has been identified as being responsible for the red coloration present in many *Heliconius* species: *optix* (Reed et al., 2011; Papa et al., 2008; Martin et al., 2014; Supple et al., 2013; Zhang et al., 2017). Optix is a highly conserved homeobox-containing transcription factor that is a member of the SIX class of proteins with well-defined function in *Drosophila* eye development (Seimiya and Gehring, 2000). It is thought that *optix* was co-opted from its role in eye development to a novel role in butterfly wing color patterning in not only *Heliconius* butterflies, but also in butterflies throughout the Nymphalid family (Zhang et al., 2017; Martin et al., 2014). Interestingly, research on this gene has also provided evidence of *optix* playing a role in scale development and structural coloration in addition to color patterning (Zhang et al., 2017; Martin et al., 2014)—suggesting that *optix* is a "master regulator" gene, or a gene capable of coordinating the development of multiple traits. Although the role of *optix* in the development of these traits is convincing, we do not yet understand how this gene gained such a pleiotropic nature.

To better understand how such developmental integration of multiple traits works, we must first understand and define the role of *optix* in each individual trait. Here, I characterize the role of *optix* in wing color patterning in the butterfly *Heliconius erato lativitta* (*H.e.lat*). To do this, I address two major questions: (1) To what extent are cis-regulatory elements, or CREs,

responsible for *optix*'s role in wing color patterning? And (2) What are the downstream targets of Optix and how do they affect color patterning? I do this by functionally validating both upstream CREs and downstream ommochrome genes using the CRISPR/Cas9 system. By investigating both the upstream regulation of *optix* as well as the downstream effect of genes it may be controlling, my work provides important insight into how this network acquired its developmental role in the evolutionarily important process of butterfly wing color patterning.

REFERENCES

Martin, A., Mcculloch, K. J., Patel, N. H., Briscoe, A. D., Gilbert, L. E., & Reed, R. D. (2014). Multiple recent co-options of Optix associated with novel traits in adaptive butterfly wing radiations. EvoDevo, 5(1), 7. doi:10.1186/2041-9139-5-7

Papa, R., Morrison, C. M., Walters, J. R., Counterman, B. A., Chen, R., Halder, G., . . . Mcmillan, W. O. (2008). Highly conserved gene order and numerous novel repetitive elements in genomic regions linked to wing pattern variation in Heliconius butterflies. BMC Genomics, 9(1), 345. doi:10.1186/1471-2164-9-345

Reed, R. D., Papa, R., Martin, A., Hines, H. M., Counterman, B. A., Pardo-Diaz, C., . . . Mcmillan, W. O. (2011). Optix Drives the Repeated Convergent Evolution of Butterfly Wing Pattern Mimicry. Science, 333(6046), 1137-1141. doi:10.1126/science.1208227

Seimiya, M. & Gehring W.J. (2000). The Drosophila homeobox gene optix is capable of inducing ectopic eyes by an eyeless-independent mechanism. Development, 127(9), 1879–1886.

Zhang, L., Mazo-Vargas, A., & Reed, R. D. (2017). Single master regulatory gene coordinates the evolution and development of butterfly color and iridescence. Proceedings of the National Academy of Sciences, 114(40), 10707-10712. doi:10.1073/pnas.1709058114

CHAPTER 2

CHARACTERIZING THE FUNCTION OF PREDICTED OPTIX CRES IN *H.E.LAT*

INTRODUCTION

optix is an important gene in creating the diverse red color patterns observed across Nymphalid butterflies, including many diversely patterned *Heliconius* species (Reed et al., 2011; Martin et al., 2014; Zhang et al., 2017). Genome-wide association studies (GWAS) conducted in *Heliconius* implicate a non-coding region adjacent to *optix* as the causative locus for pattern variation, and protein sequence comparison between *Heliconius* species reveals little to no variation in the Optix protein itself—suggesting the involvement of cis-regulatory elements, or CREs, in the divergent use of this gene (Supple et al, 2013, Reed et al., 2011). CREs are increasingly thought to be important drivers of phenotypic diversity (Wittkopp and Kalay, 2011), however there is no functional proof of these elements being the causative factors of *Heliconius* wing pattern diversity. Despite the recent advancement of genome editing technologies such as the CRISPR/Cas9 system, validating these GWAS regions is difficult considering many are multi-kilobases in size. To more narrowly define potential *optix* CREs, Dr. James Lewis obtained comparative ATACseq data sets between two distinctly patterned races of *H.erato—H.e.lativitta* and *H.e.petiverana*—and compiled a list of candidate *optix* CREs that range from about 300bp to 1000bp (Fig. 2.1). Here I describe the work I did to validate these candidate *optix* regulatory regions using the CRISPR/Cas9 system in an effort to better understand to what extent CREs are responsible for the role of *optix* in wing color patterning.

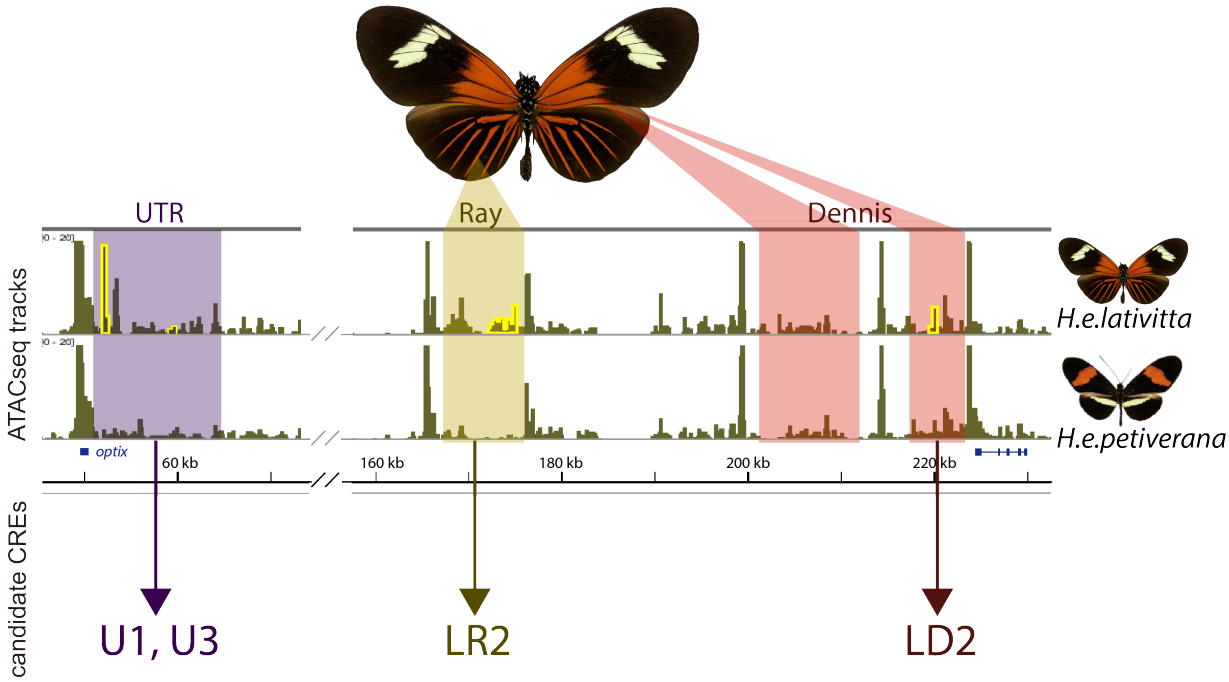

Figure 2.1 Comparative ATACseq in *H.e.lat* and *H.e.pet*. Comparative ATACseq tracks showing accessible genomic regions (represented by green peaks)—and therefore candidate CREs—near *optix*. Previously identified GWAS regions are represented by the yellow and red shaded regions. Candidate CREs investigated in this study are highlighted in yellow and labelled below, with U1 and U3 coming from an untranslated region near *optix* (purple), LR2 from within the "ray" region (yellow), and LD2 within the "dennis" region (red).

METHODS

Animals

H.e.lat pupae were obtained from Heliconius Works in Ecuador through LPS Imports and reared in a growth chamber at 27˚C with 60% humidity in a 12:12 light/dark cycle. Once eclosed, adults were released into greenhouse cages kept at 27˚C with ~80% humidity and fed pollen and nectar from *Lantana camara* and *Psiguria warscewiczii* plants. Oviposition was induced by introducing the host plant *Passiflora biflora*. Injected eggs were reared to adulthood in the growth chamber and fed fresh, young *P. biflora* clippings.

Identifying candidate optix CREs

Previously completed ATACseq data analysis (James Lewis) identified a list of *H.e.lat*-specific candidate CREs by comparing transposon-accessible genomic regions of DNA extracted from wing tissue in a rayed race (*H.e.lat*) versus a non-rayed race (*H.e.pet*) at the developmental stage where *optix* is shown to be highly expressed (Hines et al., 2012). These accessible regions were defined as candidate CREs if a peak was present in the rayed species and absent in the non-rayed species.

Cas-9 mediated genome editing

CRE knockouts were produced following the protocol of Zhang and Reed (2017); a summary of the protocol used in this study can be found below.

sgRNA Design and Production: sgRNA templates were designed by Anyi Mazo-Vargas following the $GGN_{20}NGG$ or $N_{20}NGG$ rule on either the sense or antisense DNA strand, and chosen if they displayed minimal BLAST hits to the *H.e.lat* reference genome to avoid off-target effects. Four to six sgRNAs were designed per element, with the intent to inject a total of ~4 sgRNAs per element to increase the chances of sgRNA binding in non-coding (and therefore less conserved) genomic regions. All sgRNA templates were generated by PCR amplification using a forward primer containing a T7 polymerase binding site and the sgRNA binding site and a reverse primer containing the rest of the sgRNA sequence (Bassett et al. 2013; Bassett and Liu 2014). Final sgRNAs were generated using the T7 MEGAscript Kit (Ambion, Cat. No. AM1334) followed by a phenol/chloroform extraction and isopropanol precipitation. Two groups of sgRNAs were created for this study: one group targeting the flanking regions of each element to induce large deletions of the entire peak region, and one group targeting within the peak region to induce smaller deletions. All sgRNAs used in this study are listed in Table 2.1.

Cas9: The Cas9 protein used in this study was purchased from PNABio (Cat. No. CP01) and resuspended to a final concentration of 1000ng/µl.

Embryo Injections: Eggs were collected from wild-type (WT) *H.e.lat* and lined up on double-side tape placed on a microscope slide. Borosilicate glass needles (Sutter Instrument, I.D.:0.5mm) were pulled following the specifications of Zhang and Reed (2018) and used for microinjection at an injection pressure of 10 psi. The concentrations of sgRNA:Cas9 injected for each element varied, and is summarized in Table 2.2. Adults wings were separated from the body and screened under magnification for mutant phenotypes, and the bodies were stored in -20˚C for further genotyping. Egg injections were a combined effort between myself and Anyi Mazo-Vargas.

Genotyping: Genomic DNA was extracted from adult thorax muscle tissue using an SDS-based buffer method (Chen et al., 2010). Genotyping primers, designed by both Anyi Mazo-Vargas and myself to flank the intended deletion region, are summarized in Table 2.1. Targeted regions were PCR amplified and visualized by agarose gel electrophoresis. Mutant bands (bands that were smaller or larger than the expected WT band, indicating an insertion or deletion) were gel extracted using the NEB Monarch DNA Gel Extraction Kit (NEB, Cat. No. T1020), ligated into the pGEM T-easy vector system (Promega, Cat. No. A1360), cloned into NEB DH5-alpha high-efficiency competent cells (NEB, Cat. No. C2987H) and sequenced using sanger sequencing. For individuals that did not show a mutant band, "WT" bands (bands that were the same size as expected if no insertion or deletion occurred) were gel extracted, ligated, cloned and sequenced to investigate the possibility of small indels that might not be detected by agarose gel electrophoresis. Sequences were then aligned to their respective reference sequence (downloaded from the *H.e.lat* genome available online at lepbase.org) as well as five WT

sequences that were cloned from a single WT individual, when available. Alignments were completed on the NCBI multiple sequence alignment online tool or through the ApE program.

Imaging

Butterfly wings were photographed using a Zeiss Axiocam 506 color camera with a Plan Apo S 1.03FWD 60mm Macro Lens on a SteREO DiscoveryV20 microscope.

RESULTS

Functional validation of candidate optix CREs using the CRISPR/Cas9 system

Four candidate *optix* CREs were investigated in this study. Two of these elements, LD2 and LR2, have been shown to be associated with the "dennis" and "ray" patterns of *Heliconius* butterflies, respectively (Supple et al., 2013; Van Belleghem et al., 2017) (Fig. 2.1). The remaining two elements, U1 and U3, are within an untranslated region immediately 3' of the *optix* coding region and are also associated with red patterning, though to a somewhat lesser degree when compared to LD2 and LR2 (Supple et al., 2013). Interestingly, deletion of these four elements all resulted in wing vein mutations (Figs. 2.3-6), although to varying degrees. For clarity, I will use terminology described by James A. Scott to define the vein mutations observed (Fig. 2.2) (Scott, 1986).

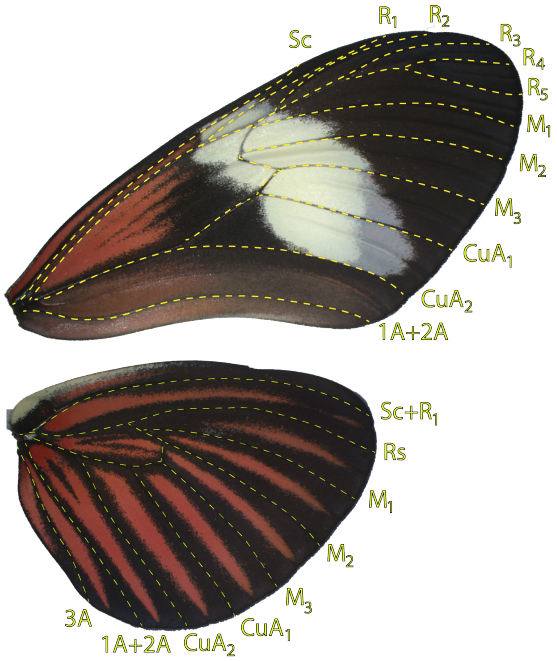

Figure 2. 2 Vein nomenclature of *H.e.lat* forewings and hindwings.

U1: The U1 element is within a genomic region that shows an association with red patterning and contains a differential ATAC peak in *H.e.lat*, suggesting its potential role in the wing patterning of this race (Fig 2.1) (Supple et al., 2013). Three clear U1 mutants were obtained following CRISPR injections: U1-29, U1-37, and U1-84 (Fig. 2.3). U1 mutations resulted in asymmetric hindwing vein abnormalities (U1-29, -37) as well as abnormal wing scale development on both the dorsal and ventral sides of the wing (U1-84). U1-29 displayed a truncation of the M_1 hindwing vein towards the distal end of both hindwings, with the proximal end of M_1 intact, resulting in the flanking ray patterns merging towards each other at the distal margin of the hindwing (Fig. 2.3A). U1-37 showed M_1 truncation on both the proximal and distal ends of the vein on both hindwings, with more vein loss occurring on the left hindwing when compared to the right, and resulting in vein merging similar to U1-29 (Fig. 2.3B). U1-84 displayed a non-vein phenotype, with an asymmetric patch of decreased scales present on both sides of one hindwing (Fig. 2.3C).

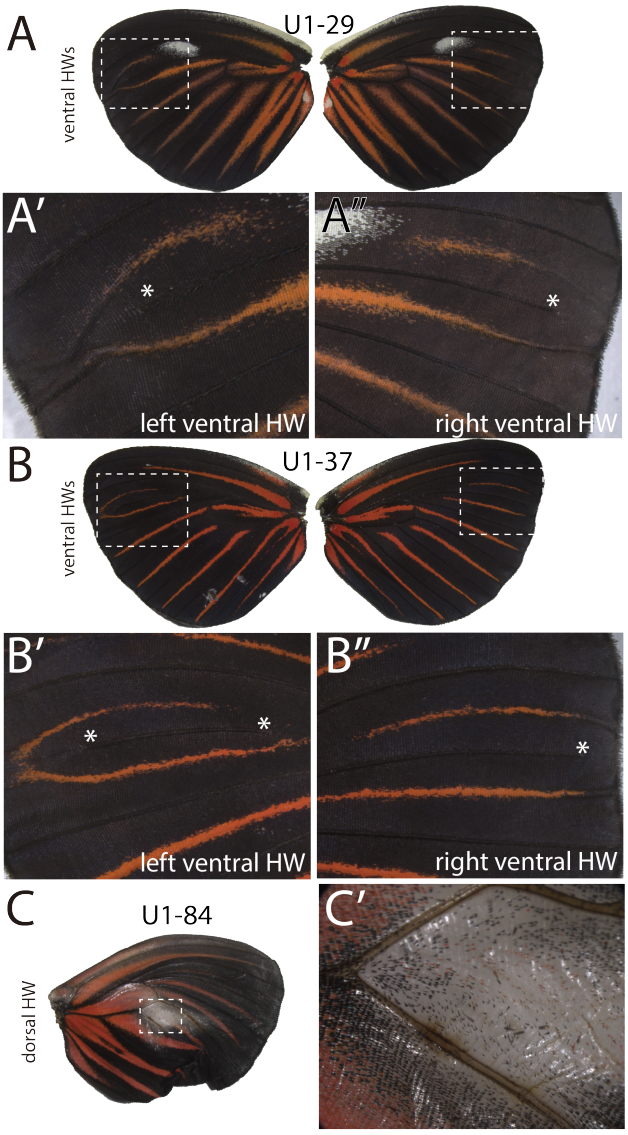

Figure 2.3 U1 knock-outs in *H.e.lat*. (A) Hindwing vein mutations seen on the U1-29 adult following CRISPR/Cas9 embryo injections. The M_1 vein on both HWs prematurely ends before reaching the most distal region of the wing, causing the flanking red rays to begin to merge. (B) Hindwing vein mutations on the U1-37 adult. On both HWs, the M_1 vein fails to reach the most distal end of the wing, and prematurely ends before reaching the most proximal region of the left HW (B'). (C) Asymmetrical scale loss seen on both the ventral and dorsal side of the U1-84 adult. Asterisks (*) mark where veins end.

U3: The U3 element also shows association with red patterning and contains an ATAC peak in *H.e.lat* (Fig 2.1) (Supple et al., 2013). U3 mutations resulted in hindwing vein abnormalities (U3-9 and U3-10) and abnormal eye pigmentation (U3-10) (Fig 2.4). The left hindwing of U3-9 showed a similar phenotype as U1-37, with M_1 truncated at both the proximal and distal ends of the vein.

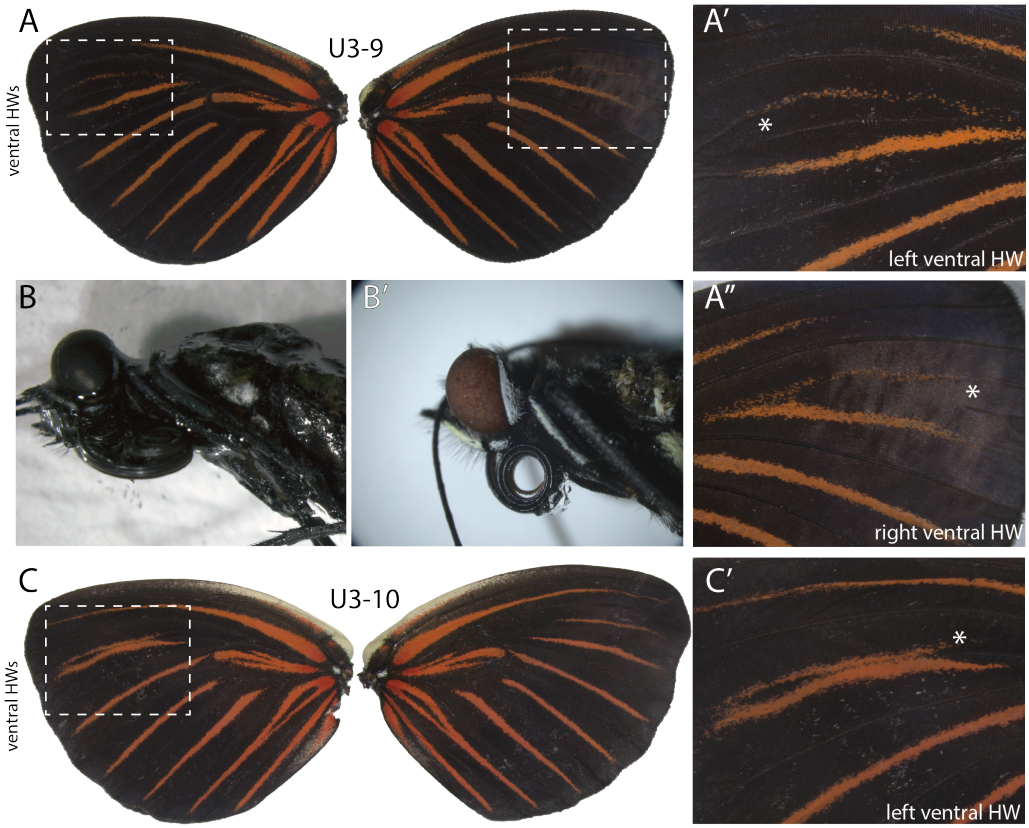

Figure 2.4 U3 knock-outs in *H.e.lat*. (A) Asymmetrical vein mutations on the U3-9 adult. On the left ventral HW, the M_1 vein fails to reach the most distal region of the wing (A'), while the M_1 vein on the right ventral HW is intact in the distal region but disappears as it moves inward towards the body (A''). (B) The red eye phenotype seen in the U3-9 adult (B') compared to WT black eyes (B). (C) Asymmetrical vein mutations on the U3-10 adult. On the left ventral HW, the M_1 vein ends near the proximal region of the wing, causing a nearly full merger of the flanking ray patterns. Asterisks (*) mark where veins end.

The right hindwing, however, showed a portion of M_1 present on the distal margin of the wing that truncated towards the proximal region of the wing, resulting in a merging of rays towards the body of the butterfly (Fig. 2.4A). U3-9 also showed a red eye phenotype, in comparison to the WT black eye coloring normally seen in these animals (Fig. 2.4B). U3-10 displayed a truncation of M_1 early in the proximal-to-distal axis of the vein, resulting in a nearly full merger of the flanking ray patterns (Fig. 2.4C).

LD2: The LD2 element is within a genomic region shown to be associated with the "dennis" pattern of *Heliconius* butterflies (Fig. 2.1), however no dennis phenotypes were observed among the 56 LD2-injected adult obtaineds. LD2 mutants also displayed vein abnormalities (LD2-35), though in different veins when compared to U1 and U3 mutants, as well as one individual showing wing scale loss (LD2-11). LD2-35 showed vein mutations in both the forewings and hindwings (Fig. 2.5A). On the forewings, the

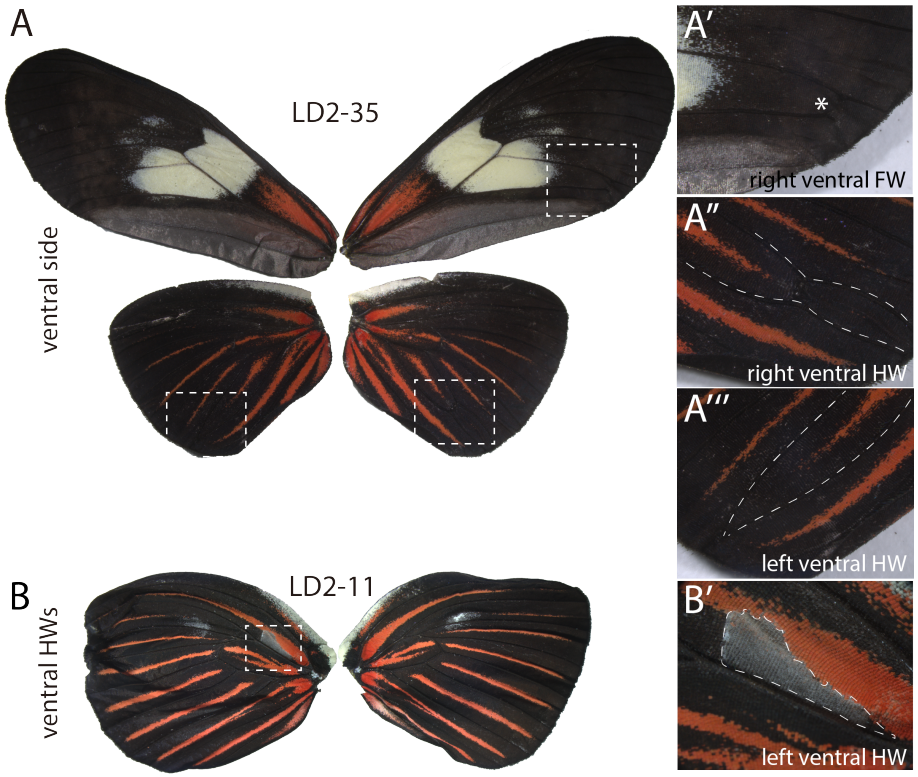

Fig. 2.5 LD2 knock-outs in *H.e.lat.* (A) LD2-35 showed three vein mutations, one on a forewing and one on each hindwing. On the HWs, the M_3 and CuA_1 veins (outlined by dotted lines) fully merge at the distal-most point of the left ventral HW and partially merge on the right ventral HW, causing a shortened ray in each case (A'' and A'''). The CuA_2 vein on the right ventral FW stops before reaching the distal region of the wing, shown by an asterisk. (B) The LD2-11 adult shows a loss of scales on the ventral side of the left HW. The scale-loss spans both red- and black-pigmented areas and is outlined in B'.

CuA_2 vein of the right ventral wing failed to develop to the distal margin. On both hindwings, M_3 and CuA_1 fully merge on the left ventral wing and partially merge on the right ventral wing at the distal margin, causing a shortened ray present between those two veins in both cases.

LD2-11 showed an asymmetric loss of scales that spans the red and black pigmented regions between the Sc/R_1 and Rs veins of the right ventral hindwing (Fig. 2.5B).

LR2: The LR2 element is within a genomic region associated with the "ray" pattern of *Helioconius* butterflies (Fig. 2.1). LR2 mutations resulted in forewing and hindwing vein abnormalities (LR2-8, LR2-49), wing-conjugation scale abnormalities (LR2-8), and a clonal loss of red pigmentation in one ventral forewing (LR2-41). LR2-8 displayed arguably the most dramatic phenotype of all the CRE mutants obtained in this study (Fig. 2.6A). On the left ventral hindwing, M_1 is completely missing—resulting in a full merging of the rays flanking the missing vein. The R_5 wing vein of the left ventral forewing is also abnormal, failing to develop to the distal end of the wing margin. And, finally, a patch of mutated wing-conjugation scales is asymmetrically present on the ventral side of the left forewing, showing a similar phenotype to *optix* KO adults (Zhang et al., 2017). LR2-49 also shows a forewing vein phenotype, with the M_1 wing vein failing to develop to the distal end of the left forewing (Fig. 2.6C). LR2-41 showed the only pigment-related phenotype obtained in this study, with a patch of decreased red pigment on the ventral side of the left forewing (Fig. 2.6B).

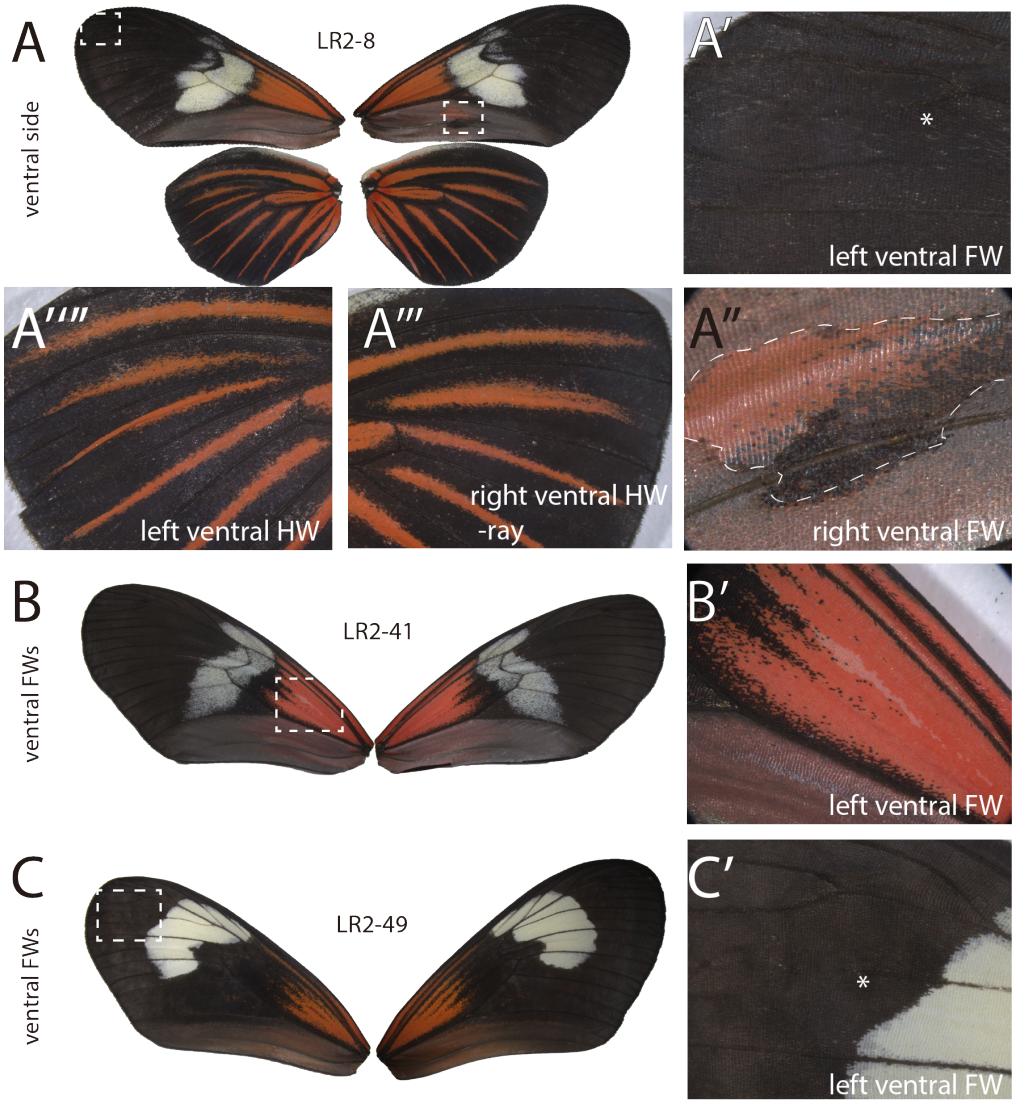

Fig. 2.6 LR2 knock-outs in *H.e.lat*. (A) The LR2-8 adult showed early truncation of the R_5 vein on the left ventral FW (A'), morphological transformation of wing conjugation scales (A''), and complete loss of the M_1 vein of the right ventral HW (A'''). (B) LR2-41 showing a relatively small, asymmetrical clone of decreased pigmentation in the "dennis" element of the left ventral FW. (C) LR2-49 with an asymmetrical mutation of the M_1 FW vein on the left side. Asterisks (*) mark where veins end.

Confirming optix CRE KOs through genotyping

Confirming successful CRE targeting and deletion through the more commonly used genotyping by PCR method proved difficult for many of the mutants obtained in this study. Although all of the mutant individuals displayed a clear band of the expected WT size following

PCR amplification of the genomic region containing the targeted element (Figs. 2.7-10), most reactions failed to amplify a mutant band (i.e., a smaller or larger band indicating a deletion or insertion event) as expected (Zhang et al., 2017). For those individuals, "WT" bands were gel extracted, cloned, and sequenced to investigate the possibility of small indel mutations that may nothave been detectable through agarose gel electrophoresis. These sequences were then compared to several true WT sequences from uninjected individuals for comparison, when available. A summary of the sequencing results for each mutant discussed in this study can be found in Table 2.3.

U1: Genomic DNA was extracted from each of the three U1 mutants, PCR amplified, gel extracted, cloned, sequenced, and aligned to the *H.e.lat* reference genome. U1 sequences from an uninjected individual were also PCR amplified, cloned, and sequenced and provided five additional WT sequences for further comparison.

U1-29: PCR amplification of the U1 region in this individual failed to produce a mutant band (Fig. 2.7B). "WT" clones were sequenced and investigated for any unique indels or deletion/insertion events. Of the three clones sequenced, one showed sequence variation unique to the injected individual when compared to WT individuals (Table 2.3). These unique indels, however, were not near any expected sgRNA cut sites and are therefore difficult to interpret as naturally occurring sequence variation or as a result of intended Cas9-mediated genome editing.

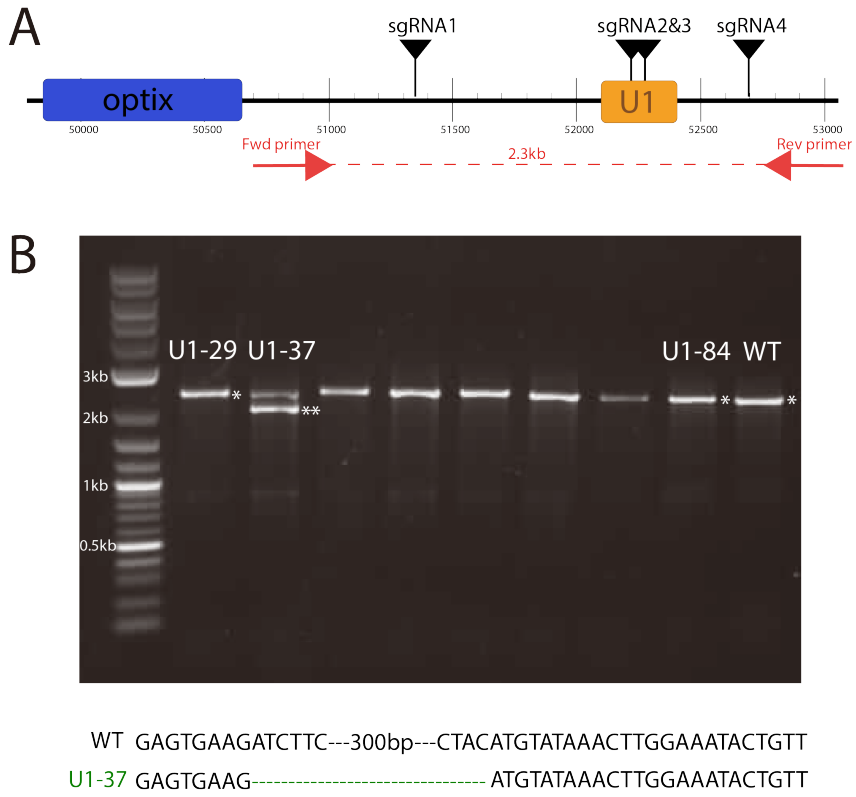

Figure 2.7 CRISPR/Cas9 and genotyping experimental plan for the U1 element. (A) Diagram of the sgRNA target sites and genotyping primers for the U1 candidate CRE. Genotyping primers are represented by red arrows and the amplified region is marked in red. (B) Genotyping by PCR results for the 3 U1 mutants discussed in this paper: U1-29, U1-37, U1-84 (bands not named are irrelevant to this study). Asterisks mark the bands that were gel extracted, cloned, and sequenced. Sequencing results for the U1-37 "mutant" band (**) are below, showing a 300+bp deletion in the amplified region.

U1-37: This individual was one of the few cases where genotyping by PCR was successful in identifying a CRE KO event. PCR amplification resulted in a clear mutant band (Fig. 2.7B) and subsequent sequencing of this band revealed a 312bp deletion that included an entire sgRNA binding site (Table 2.3).

U1-84: PCR amplification of the U1 region in this individual showed no mutant band (Fig. 2.7B) and of the four "WT"-sized clones that were sequenced, two showed minor indels (<6bp) in three areas of the element (Table 2.3). None of these indels were near

an expected sgRNA cut site, and are therefore difficult to connect to directed Cas9-mediated genome editing.

U3: Of the two U3 mutants obtained, genomic DNA was only extracted from U3-9 because the U3-10 sample was lost during a freezer cleanout. PCR amplification of the U3 region in U3-9 showed no mutant band (Fig. 2.8B). 10 clones were sequenced from the extracted "WT"-sized band but showed no sequence difference when compared to the WT sequence, outside of a few expected SNPs (data not shown).

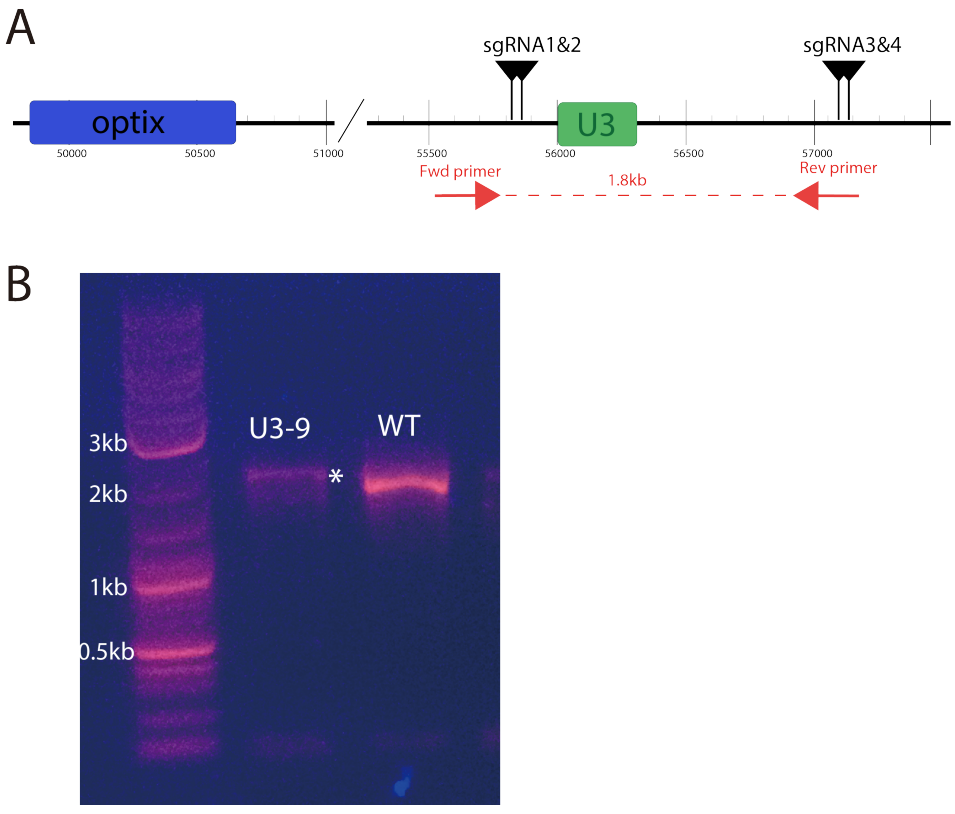

Figure 2.8 CRISPR/Cas9 and genotyping experimental plan for the U3 element. (A) Diagram of the sgRNA target sites and genotyping primers for the U3 candidate CRE. Genotyping primers are represented by red arrows and the amplified region is marked in red. (B) Genotyping by PCR results for U3-9, showing the band that was gel extracted, cloned, and sequenced with an asterisk.

LD2: Genomic DNA was extracted from both of the LD2 mutants obtained and PCR amplified (Fig. 2.9). Five wild-type clones from an uninjected, WT individual were also sequenced and used for reference.

LD2-11: For this individual, the LD2 region was PCR amplified and the "WT" band was gel extracted and immediately sent to sequencing (Fig. 2.9B). This resulted in only a portion of the amplified region being sequenced; however, the fragment that was sequenced showed a 15bp insertion at an expected cut site—providing strong evidence for the observed phenotype being a result of the intended genomic alteration (Table 2.3).

LD2-35: PCR amplification of the LD2 region in this individual showed no mutant band (Fig. 2.9A). Six "WT"-sized bands were cloned, sequenced, and aligned to the *H.e.lat* genome as well as the 5 true WT sequences obtained from the uninjected individual. Three of the six LD2-35 sequences showed a 91bp deletion between two sgRNA binding sites that was not present in any of the WT sequences investigated (Table 2.3). Although the deletion was not near an expected cut site, its exclusive occurrence in only the injected animal provides support for some effect of Cas9-mediated editing.

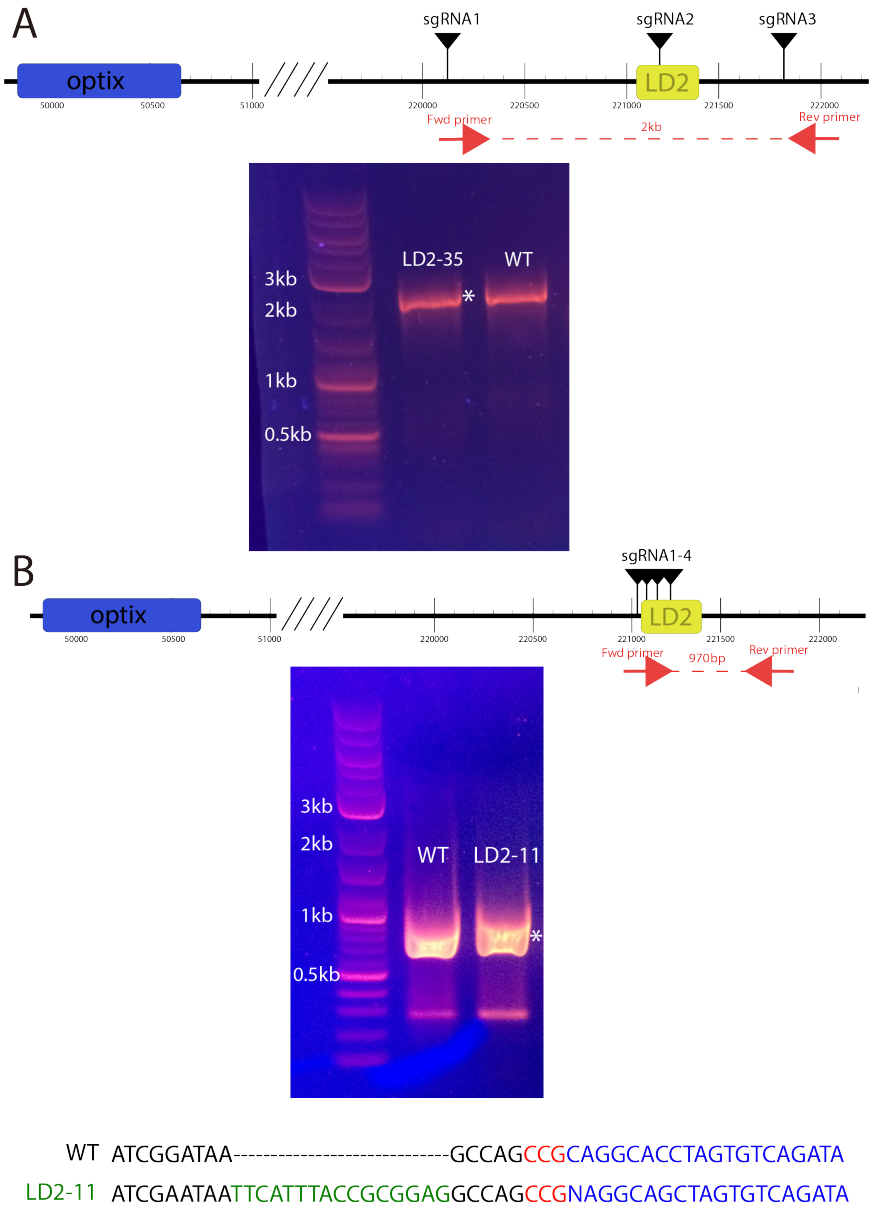

Figure 2.9 CRISPR/Cas9 and genotyping experimental plan for the LD2 element. (A) A diagram showing the experimental plan that produced LD2-35, with sgRNAs flanking the LD2 region. Genotyping primers are represented by the red arrows and the PCR amplified region is shown in red. The PCR results are shown below, with an asterisk showing the band that was gel extracted, cloned, and sequenced. (B) A diagram showing the experimental plan that produced LD2-11, with sgRNAs designed to target within the LD2 element. Again, genotyping primers are represented by red arrows with the amplified region shown in red. An Asterisk marks the gel extracted band that was sent to sequencing, with the sequencing results shown below. The blue text shows an sgRNA binding site, with the red representing the PAM sequence.

LR2: The LR2 region was the most difficult to genotype for several reasons. It was the largest of the candidate CREs at 1000bp, contained poor GC-content making for difficult primer and sgRNA design, and was highly repetitive. Because of these challenges, the PCR products from these experiments often showed misamplification. Furthermore, the intended PCR product was close to 3kb in size and proved difficult to ligate. This resulted in many failed cloning attempts and a failure in obtaining WT sequences for comparison. All sequencing results were therefore aligned with only the reference genome for this element.

LR2-8: For reasons mentioned above, amplification of the LR2 region often resulted in misamplification events (Fig. 2.10B). Though the PCR reaction yielded a potential mutant band in the agarose gel, the one clone sequenced from this band was simply the last 1kb portion of the intended amplified region, beginning 676bp away from the nearest expected cut site, with no sequence differences from the reference genome. It is therefore difficult to conclude whether or not this sequence is a result of intentional genome editing or simply sample degradation.

LR2-41: PCR amplification of the LR2 region in this individual yielded similar misamplification events (Fig. 2.10B). Additionally, when attempting to clone and sequence "WT" bands from this individual, most of the cloning attempts failed. Of the three clones that I was able to obtain, only ~500bp of the 2.8kb insert was ligated into the cloning vector. Of those 500bp that were sequenced, no mutations were found.

LR2-49: PCR amplification of this individual also yielded poor results (Fig. 2.10B). The LR2 "WT"-sized region was similarly difficult to clone, resulting in only one successful cloning event of a partial ligation (~650 of the 2800bp). There were no sequence differences observed in the obtained sequence when compared to the WT reference.

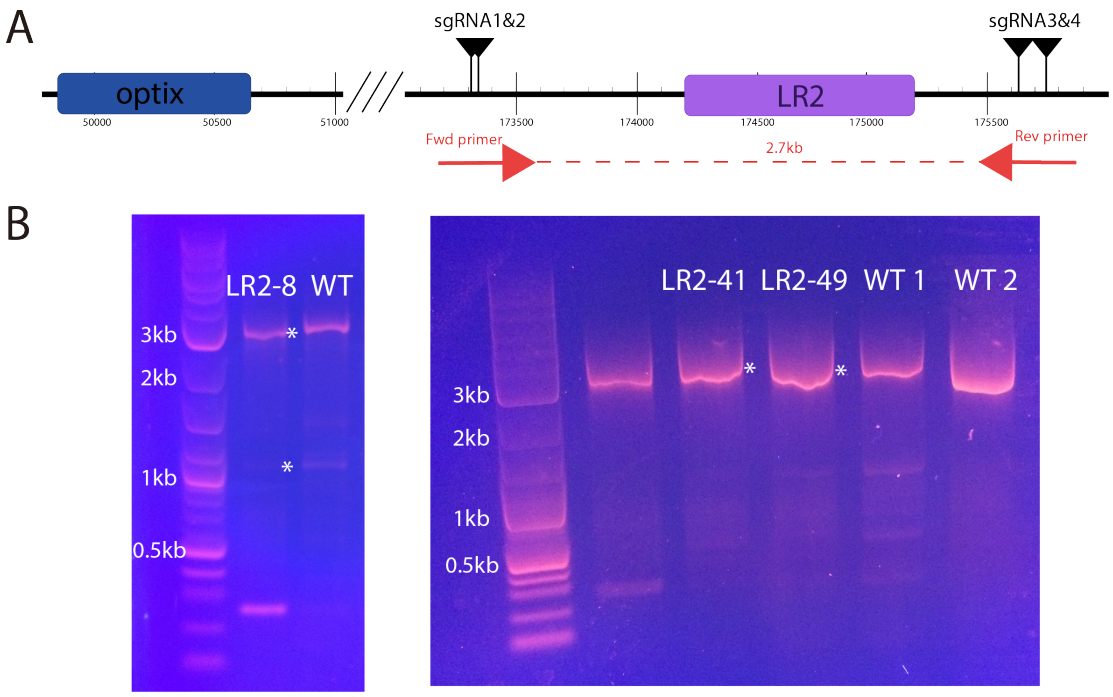

Figure 2.10 CRISPR/Cas9 and genotyping experimental plan for the LR2 element. (A) Diagram of the sgRNA target sites and genotyping primers for the LR2 candidate CRE. Genotyping primers are represented by red arrows and the amplified region is marked in red. (B) Genotyping by PCR results for the mutants of interest: LR2-8, -41, and -49. Asterisks mark the bands that were gel extracted, cloned, and sequenced.

DISCUSSION

Previously published research has shown that *optix* has a role in red pigmentation patterning as well as in the morphological development of wing conjugation scales in *H. erato* (Zhang et al., 2017). Following these findings, I expected *optix* CRE KOs to produce similar, albeit less extreme, versions of the -*optix* phenotypes in a "clonal" or "mosaic" fashion—meaning the adults will show a combination of wild-type and abnormal patterning due to the incomplete incorporation of the CRISPR/Cas9 complex in early cells leading to a mix of wild-type and mutant cells making up the adult animal. The mutants observed in this study, however, showed a variety of phenotypes ranging from the expected to the unexpected. Two of the ten mutants displayed expected phenotypes: one with a patch of decreased red pigmentation (LR2-41) and another showed a morphological change in scale structure in a patch of wing conjugation scales

(LR2-8). Additionally, two mutants (LD2-11 and U1-84) showed complete loss of wing scales on either the ventral or both the ventral and dorsal side of the hindwing, respectively. Although this complete loss of scales was not observed in the previously published *optix* KO results, it is not altogether surprising considering the predicted involvement of this gene in scale morphology (Zhang et al., 2017). What was surprising, however, was the wing vein phenotypes observed in the majority of mutants obtained in this study. Seven of the ten mutants showed abnormalities in wing vein development—some with veins terminating before reaching the distal regions of the wings, others with veins that disappeared towards the proximal region of the wings, and one with the complete loss of a hindwing vein. While these results were initially surprising, recent research has shed light on a previously undescribed function of *optix*: positioning the longitudinal wing veins in *Drosophila melanogaster* (Martín et al., 2017). These studies report Optix as being necessary for L2 vein development on the anterior portion of the *Drosophila* wing. Furthermore, they provide a possible regulatory mechanism for this role of Optix, describing Decapentaplegic acting as an activator and Brinker and Spalt acting as repressors. Further research will need to be done in order to confirm these interactions in *Drosophila* and whether they occur in *Heliconius*; however, this model provides an interesting framework to explore the potential role of Optix in butterfly wing vein patterning.

Another interesting outcome of this study was the low yield of animals with wing abnormalities despite the proven success of using CRISPR in multiple butterfly species, including *Heliconius* (Zhang et al., 2017; Zhang and Reed, 2017). Out of the 254 CRISPR-injected adults obtained, only ten displayed a clear mutant phenotype. This low frequency could be the result of a few different factors. First, the regions targeted in this study were intergenic and therefore less likely to be conserved. This variability may have affected sgRNA binding and, subsequently, lowered the likelihood of producing the intended double-stranded cuts. To explore this possibility, I looked at SNP data available from 4 individuals and investigated sequence variation in each of

the sgRNA binding sites (data not shown). Although most sgRNA binding sites showed little to no variation, some SNPs were present and previous research has shown even a single base-pair change in an sgRNA binding site can affect binding efficiency (Lessard et al., 2017).

Yet another explanation for the low yield of mutant phenotypes could relate to the idea of enhancer redundancy. Multiple, discrete enhancer regions for a single gene have been identified in many systems and are hypothesized, and in some cases proven, to provide phenotypic robustness in their redundancy—with many studies showing minor phenotypic mutations following the loss of a single redundant enhancer, similar to the findings in this study (Frankel et al., 2010; Lam et al., 2015; Osterwalder et al., 2018). This redundancy is thought to be evolutionarily useful for buffering against genetic and environmental variability. Considering the role of *optix* in multiple important developmental pathways, enhancer redundancy seems likely in the regulation of this gene; however, more research needs to be done in order to confirm this hypothesis.

The difficulty in genotyping the animals obtained in this study unfortunately prevents any solid conclusions to be made regarding the connection between the candidate CREs and the phenotypes observed. There are, however, a few possible methods that may circumvent the PCR issues encountered here in future studies. First, the source of genomic DNA amplified was not wing tissue, but rather thorax tissue—a part of the body that showed no observable phenotypes and therefore may only be composed of wild-type cells. Thorax tissue was used over wing tissue because of the difficulty in extracting DNA from wings. In order to get DNA from the wing tissue, you must immediately ruin the wings upon eclosing, which prevents any thorough investigation of potential mutant phenotypes. Because thorax tissue previously showed successful CRISPR genome editing in *Heliconius,* I expected similar results in this study—however, that choice may have prevented the genotyping confirmation of these CRE

mutants (Zhang et al., 2017). In future work, it may be advisable to extract DNA from the wings rather than from the thorax. Additionally, alternate genotyping methods may need to be considered. The frequency of editing events may be significantly low, making PCR amplification inadequate in confirming mutations. For this reason, high-throughput sequencing methods such as MiSeq should be considered for future attempts.

TABLES

Table 2.1: sgRNAs and primers used in this study

Target	sgRNA Name	sgRNA sequence	Genotyping Primers
U1	29	CCGCTCTACTTTACTTTATT	F: CGAGCAGTCCGTGATGTGAT
	31	ATCTTGTAAGTTACATGGTC	R: CGATGCGCTAAGTGTTTCGT
	u1p1	GGCTCGGTGCAATTAATAAA	
	u1p2	GTATTATTCAGGACATGGCT	
	u1p3	ACAGCTTGGTCCCGCTCCAG	
	u1p4	ATAGGATCTACTGGAGCCAC	
U3	36	AACTTTTACATTCTACACAT	F: CAGTTACCCAGTGTTCCACGA
	37	ACGCACACGTCAGGGTAAGC	R: CGTAATGTGTGGCTTGCACGTA
	38	CCGTGGATGTGATAAGTTCT	
	39	ATGTGAGGAGGATGATAGTC	
	u3p1	GCAGCGTATATCAGGAGGTG	
	u3p2	ACAATAACGCAGCGTATATC	
	u3p3	ATTTGGCACAAAGAGCAATT	
	u3p4	GGAACGTAGTATAAGTTCCC	
LD2	11	AACGTGGTTACCGGCCGACG	F1: TCTTCTGAAAGGACGCCAAAC
	13	CAGGGAAGCGATTGTCATGT	R1: TGTCCGTGTCGTCGCTATTT
	ld2p1	GCTGGCTTATCCGATTCAAG	
	ld2p2	TATCTGACACTAGGTGCCTG	F2: TGGCCTCTATTGTGCTACGTG
	ld2p3	GTTGCGATGTACGATATCAC	R2: GGTTCTCAGGTACTTCCTAGGGT
	ld2p4	TCGTCTCACGGGCGGGGCGT	
LR2	22	AAAAGGGACCGAATCTATGA	F: CGCCGTTGCAGTAGCTCTAT
	23	ACCGACCCTGAACTGAACGC	R: CGAACCACAGTCACACAACCAG
	25	CCACTAAAATTAACATCACA	
	26	GCAACTATTAATAATATAAA	
	lr2p1	TGTGCCCCCAAACTGCGCGG	
	lr2p2	CTTTATTCAATCACGGCGGT	

Table 2.2: Injection Summary

*****_italicized injection entries were performed by Anyi Mazo-Vargas_

ELEMENT	sgRNAS USED (TOTAL)	[sgRNA(EACH):CAS9] ng/µl	INJECTED EGGS	HATCHED EGGS	HATCH RATIO	TOTAL ADULTS	MUTANTS	MUTANT NAME
U1	29, 31, u1p1, u1p4 (4)	97:388	289	180	62%	95	3	U1-29, U1-37, U1-84
U3	*36, 37, 38, 39 (4)*	*100:400*	*79*	*NA*	*NA*	*8*	*2*	*U3-9, U3-10*
	u3p1-4 (4)	96:384	173	35	20%	11	0	
	u3p1, u3p3 (2)	192:384	51	12	24%	6	0	
LD2	ld2p1-4 (4)	96:384	96	25	26%	12	1	LD2-11
	ld2p1-4 (4)	50:200	135	46	34%	3	0	
	ld2p1, ld2p4 (2)	75:200	51	33	65%	10	0	
	11,13,ld2p4 (3)	185:486	435	249	57%	31	1	LD2-35
LR2	*22, 23, 25, 26 (4)*	*100:400*	*60*	*NA*	*NA*	*11*	*1*	*LR2-8*
	lr2p1-6 (6)	62:375	141	59	42%	15	0	
	22, 23, 25, 26 (4)	96:385	83	44	53%	21	0	
	22, 23, 25, 26 (4)	185:710	118	56	47%	31	2	LR2-41, LR2-49

Table 2.3: Sequencing Results

NAME	PHENTOTYPE	BAND SEQ	# CLONES SEQ	DETAILED SEQ RESULTS
U1-29	ventral HW veins (both)	wt	3	-5bp chr18: 51529-33; -2bp chr18: 51626-27; AAA->--T chr18: 51586-88; -3bp chr18: 52101-03; AAAACTACT->-GAA--AC- chr18: 52845-53
U1-37	ventral HW veins (both)	mutant	5	-312bp chr18: 52588-52900; +3bp chr18: 52301-02
U1-84	dorsal and ventral HW scale loss	wt	4	+5bp chr18: 50882-83; -5bp chr18: 51528-32; -4bp chr18: 52157-60
U3-9	ventral HW veins (both); red eyes	wt	10	no variation
U3-10	ventral HW veins (left)	-	-	-
LD2-11	ventral scale loss (left)	wt	none (GE prod)	+15bp chr18: 21066-67
LD2-35	ventral FW veins (right); ventral HW veins (both)	wt	6	+4bp chr18: 220575-76; -91bp chr18: 220621-220712
LR2-8	ventral FW WGS (right); ventral FW veins (left); ventral HW veins (right)	mutant	5	
LR2-41	ventral FW decreased pigmentation (left)	wt	3	no variation
LR2-49	ventral FW vein (left)	wt	1	no variation

REFERENCES

Bassett, A., & Liu, J. (2014). CRISPR/Cas9 mediated genome engineering in Drosophila. Methods, 69(2), 128-136. doi:10.1016/j.ymeth.2014.02.019

Bassett, A. R., Tibbit, C., Ponting, C. P., & Liu, J.-L. (2013). Highly Efficient Targeted Mutagenesis of Drosophila with the CRISPR/Cas9 System. Cell Reports, 4(1), 220–228. http://doi.org/10.1016/j.celrep.2013.06.020

Frankel, N., Davis, G. K., Vargas, D., Wang, S., Payre, F., & Stern, D. L. (2010). Phenotypic robustness conferred by apparently redundant transcriptional enhancers. Nature, 466(7305), 490-493. doi:10.1038/nature09158

Lam, D. D., Souza, F. S., Nasif, S., Yamashita, M., López-Leal, R., Otero-Corchon, V., . . . Low, M. J. (2015). Partially Redundant Enhancers Cooperatively Maintain Mammalian Pomc Expression Above a Critical Functional Threshold. PLOS Genetics, 11(2). doi:10.1371/journal.pgen.1004935

Lessard, S., Francioli, L., Alfoldi, J., Tardif, J., Ellinor, P. T., Macarthur, D. G., . . . Canver, M. C. (2017). Human genetic variation alters CRISPR-Cas9 on- and off-targeting specificity at therapeutically implicated loci. Proceedings of the National Academy of Sciences, 114(52). doi:10.1073/pnas.1714640114

Martin, A., Mcculloch, K. J., Patel, N. H., Briscoe, A. D., Gilbert, L. E., & Reed, R. D. (2014). Multiple recent co-options of Optix associated with novel traits in adaptive butterfly wing radiations. EvoDevo, 5(1), 7. doi:10.1186/2041-9139-5-7

Martín, M., Ostalé, C. M., & Celis, J. F. (2017). Patterning of the Drosophila L2 vein is driven by regulatory interactions between region-specific transcription factors expressed in response to Dpp signalling. Development, 144(17), 3168-3176. doi:10.1242/dev.143461

Osterwalder, M., Barozzi, I., Tissières, V., Fukuda-Yuzawa, Y., Mannion, B. J., Afzal, S. Y., . . . Pennacchio, L. A. (2018). Enhancer redundancy provides phenotypic robustness in mammalian development. Nature, 554(7691), 239-243. doi:10.1038/nature25461

Reed, R. D., Papa, R., Martin, A., Hines, H. M., Counterman, B. A., Pardo-Diaz, C., . . . Mcmillan, W. O. (2011). Optix Drives the Repeated Convergent Evolution of Butterfly Wing Pattern Mimicry. Science, 333(6046), 1137-1141. doi:10.1126/science.1208227

Scott, J. A. (1986) The butterflies of North America. Stanford University Press, Stanford, CA.

Supple, M. A., Hines, H. M., Dasmahapatra, K. K., Lewis, J. J., Nielsen, D. M., Lavoie, C., . . . Counterman, B. A. (2013). Genomic architecture of adaptive color pattern divergence and convergence in Heliconius butterflies. Genome Research, 23(8), 1248-1257. doi:10.1101/gr.150615.112

Van Belleghem, S. M., Rastas, P., Papanicolaou, A., Martin, S. H., Arias, C. F., Supple, M. A., . . . Papa, R. (2017). Complex modular architecture around a simple toolkit of wing pattern genes. Nature Ecology & Evolution, 1(3), 0052. doi:10.1038/s41559-016-0052

Wittkopp, P. J., & Kalay, G. (2011). Cis-regulatory elements: Molecular mechanisms and evolutionary processes underlying divergence. Nature Reviews Genetics, 13(1), 59-69. doi:10.1038/nrg3095

Zhang, L., Mazo-Vargas, A., & Reed, R. D. (2017). Single master regulatory gene coordinates the evolution and development of butterfly color and iridescence. Proceedings of the National Academy of Sciences, 114(40), 10707-10712. doi:10.1073/pnas.1709058114

Zhang L., Reed R.D. (2017) A Practical Guide to CRISPR/Cas9 Genome Editing in Lepidoptera. In: Sekimura T., Nijhout H. (eds) Diversity and Evolution of Butterfly Wing Patterns. Springer, Singapore

CHAPTER 3

FUNCTIONAL VALIDATION OF CANDIDATE OMMOCHROME GENES IN *H.E.LAT*

INTRODUCTION

The diversity of butterfly wing colors and patterns has been of interest to evolutionary biologists for many years due to their adaptive function, providing a link between development and evolution. Following this interest, four major groups of butterfly wing pigments have been identified: pterins, flavonoids, melanins, and ommochromes (Nijhout 1991). *Heliconius* pigmentation, specifically, is thought to be a result of two of these pigment types—melanins (tans, browns, blacks) and ommochromes (reds, oranges)—with a third molecule, 3-hydroxykyurenine (or 3-OHK), being responsible for yellow coloring (Gilbert et al., 1988). 3-OHK is not defined as one of the four major groups of butterfly pigments because it is thought to be a precursor to ommochrome pigments, and therefore likely shares the same biosynthetic pathway (Linzen 1974). In this study, I sought to identify genes involved in the butterfly wing ommochrome pathway in *Heliconius* butterflies for a few key reasons. First, ommochromes are commonly used in insect eye pigmentation (Linzen 1974) and these eye-specific ommochrome synthesis pathways are fairly well defined in various insects, including *Drosophila melanogaster*, thus providing an initial framework to explore previously identified candidate genes in the butterfly system. Second, although ommochrome pathways are relatively well understood in terms of insect eye pigmentation, their role in wing pigment synthesis has yet to be resolved. Interestingly, the use of ommochromes in butterfly wing pigmentation can be traced to a single origin in the Nymphalidae butterfly family (Nijhout 1991). This presents a great opportunity to explore the mechanisms of how biosynthetic pathways may evolve. Here I present the work I did to functionally validate candidate ommochrome genes in *H.e.lat* using the CRISPR/Cas9 system.

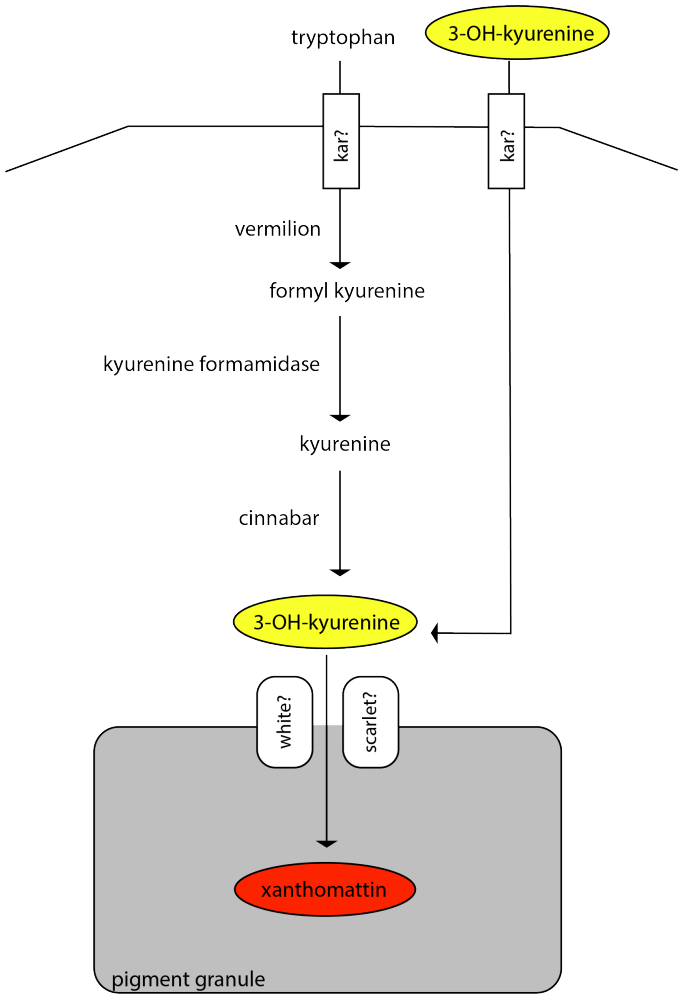

Figure 3.1 The predicted ommochrome pathway in butterfly wing pigmentation. Tryptophan serves as a precursor to the ommochrome pathway and is predicted to be imported into the cell by the karmoisin transporter. A series of enzymatic steps catalyzes the transformation of tryptophan into 3-hydroxykyurenine (3-OHK), which is then predicted to be imported into the pigment granule by a heterodimer white/scarlet transporter. The synthesis of 3-OHK into xanthomattin (the red pigment) is not yet understood. 3-OHK is also thought to be taken up from the haemolymph and is responsible for yellow coloring. This figure has been adapted from Reed et al. (2008) and Jiggins (2018).

METHODS

Animals

Animals were obtained and raised as described in Chapter 2.

Identifying candidate ommochrome genes

Candidate ommochrome genes were identified through a combination of previously completed comparative transcriptomic work in various Nymphalid butterflies as well as work done in the *Drosophila* system.

cinnabar (cn): cn was chosen for its known role in the *Drosophila* eye pigmentation pathway (Wald 1946), as well as for its gene expression patterns in both *Vanessa* and *Heliconius* butterflies. *cn* was shown to be associated with the forewing band of *H.erato* (Reed et al., 2008) and exhibited increased gene expression levels during the developmental stage where ommochrome synthesis occurs in *H.melpomene* (Ferguson and Jiggins, 2009) as well as in *Vanessa cardui* (Zhang et al., 2017). Furthermore, *cn* shows upregulation in red-pigmented *V.cardui* tissue when compared with black-pigmented tissue.

major facilitator superfamily 2 (MFS2): MFS2 was chosen for this study due to its increased expression pattern in ommochrome stage wing tissue as well as in red pigmented wing tissue in *Vanessa cardui* (Zhang et al., 2017).

Cas-9 mediated genome editing

In general, CRE knockouts were produced following the protocol of Zhang and Reed (2017).

sgRNA Design and Production: Two sgRNAs for each gene were designed with the general desire to target exons (to increase the likelihood of sequence conservation and subsequently

sgRNA binding) and to create a deletion of at least 100bp in length. sgRNA templates were designed following the N_{20}NGG rule on either the sense or antisense DNA strand and chosen if they displayed minimal BLAST hits to the *H.e.lat* reference genome to avoid off-target effects. Final sgRNAs were synthetically generated by the Synthego company and injected with Cas9 protein in RNP complex form. All sgRNAs used in this study are listed in Table 3.1.

Cas9: The Cas9 protein used in this study was purchased from PNABio (Cat. No. CP01) and resuspended to a final concentration of 1000ng/µl.

Embryo Injections: Eggs were collected from WT *H.e.lat* and lined up on double-side tape placed on a microscope slide. Borosilicate glass needles (Sutter Instrument, I.D.:0.5mm) were pulled following the specifications of Zhang and Reed (2017) and used for microinjection at an injection pressure of 10 psi. Two sgRNAs were used for both target genes. The concentration of sgRNA:Cas9 injected for *cn* was 125:250 ng/µl and for *MFS2* was 166:333 ng/µl. Injection summaries are provided in Table 3.2. Adult wings were separated from the body and screened under magnification for mutations, and the bodies were stored in -20˚C for further DNA extractions and genotyping.

Genotyping: Genomic DNA was extracted from adult thorax muscle tissue using either an SDS-based buffer method (Chen et al., 2010) or a proteinase-k digestion method (Bassett et al., 2013). DNA fragments including the intended deletion site were PCR amplified using primers flanking the sgRNA binding sites, and mutant PCR fragments were detected using agarose gel electrophoresis. Potential mutant and WT bands were gel extracted using the NEB Monarch DNA Gel Extraction Kit (NEB, Cat. No. T1020), ligated into the pGEM T-easy vector system (Promega, Cat. No. A1360), cloned into NEB DH5-alpha high-efficiency competent cells

(NEB, Cat. No. C2987H) and sequenced using Sanger sequencing. All genotyping primers used in this study are listed in Table 3.1.

Imaging

Butterfly wings were photographed using a Zeiss Axiocam 506 color camera with a Plan Apo S 1.03FWD 60mm Macro Lens on a SteREO DiscoveryV20 microscope.

RESULTS

Functional validation of two candidate ommochrome pathway genes

For this study I focused on validating the function of two candidate ommochrome pathway genes, *cn* and *MFS2*, in an effort to better understand their potential role in ommochrome synthesis as it relates to butterfly wing color patterning. These genes were chosen due to their ability to fulfill two categories of candidate ommochrome genes: those with known function in insect eye ommochrome synthesis (*cn*) and those that were previously unknown to participate in ommochrome synthesis but show increased gene expression during important pigmentation developmental stages in various butterflies (*MFS2*).

cn: Despite the strong support for the participation of *cn* in ommochrome synthesis in butterfly wings, I was unable to functionally confirm this gene's involvement in wing pigmentation in this study. Hatch ratios of *cn*-injected eggs were similar to expected levels (43.9%) and yielded 16 eclosed adults, none of which showed wing-pigmentation abnormalities. This result could reflect lack of effect of *cn* on pigmentation, or a technical issue—lack of generation of mutations with the sgRNAs. To investigate whether or not the sgRNAs were inducing the intended deletions, genomic DNA was extracted from 15 of the adults using the proteinase-K digestion method. Because this method produces a crude DNA product, PCR amplification

resulted in varied success (Fig. 3.2).

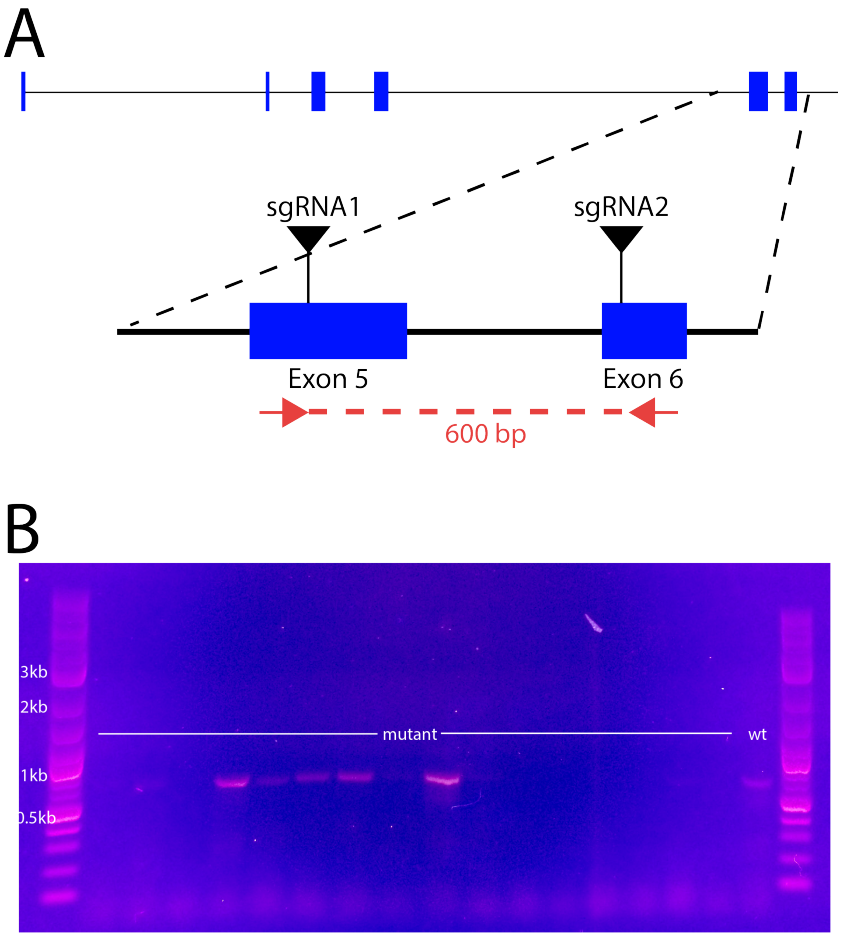

Figure 3.2 Experimental plan and genotyping results for *cn*. (A) Diagram showing the location of sgRNA binding sites in the *cn* coding region. Exons are represented by blue boxes and genotyping primers are represented by red arrows. The PCR amplified region for genotyping is highlighted in red. (B) The PCR products of 15 *cn*-injected adults (the lanes labeled "mutant") following amplification of the CRISPR-targeted region, showing varied success in amplification and inconclusive results.

Of the 15 individuals amplified, 5 showed clear amplification, 3 showed low amplification, and 7 showed little to no amplification. Of the samples that successfully amplified, none displayed a clear mutant band.

MFS2: Interestingly, multiple major facilitator superfamily (MFS) transcripts showed significant upregulation during pigment development in *V. cardui* wing tissue (Zhang et al., 2017). Of those seven upregulated MFS transcripts, four showed specific association with red regions of the wing tissue—including *MFS2*. Following preliminary CRISPR work done by Linlin Zhang and Benjamin Brock that showed loss of red pigmentation in *MFS2*-targeted *V. cardui* butterflies, I sought to validate the potential involvement of this gene in *Heliconius*

butterfly wing color patterning. Similar to the findings in *V. cardui*, *MFS2* mutant *Heliconius* butterflies injected with sgRNAs against *MFS2* and with Cas9 displayed varying degrees of strong, clonal loss of red pigmentation with no effect on black pigmentation (Fig. 3.3). Many of these butterflies showed a red-to-pink transition with altered scale structure in the mutant/pink clonal regions (Fig. 3.3A,B). Interestingly, many mutant clones extended out past the red-pigmented "dennis" element of the forewing and displayed a yellow-to-white color transition in the "band" element of the forewing, without affecting the black melanic region in between (Fig. 3.3A,C). The scales within the yellow-to-white mutant clones, unlike the red-to-pink clones, showed no morphological change in scale structure. In sum, these results suggest *MFS2* plays a role in ommochrome synthesis and, potentially, scale morphology. Unfortunately, experiments using PCR to attempt to genotype adults with abnormal wing patterns were inconclusive. Despite designing multiple genotyping primers that when BLASTed against the *H.e.lat* genome showed only a single hit, PCR amplification of the targeted region yielded multiple misamplification events (Fig. 3.4B) and sequencing results were not able to prove the existence of intended editing events in the targeted *MFS2* region (data not shown).

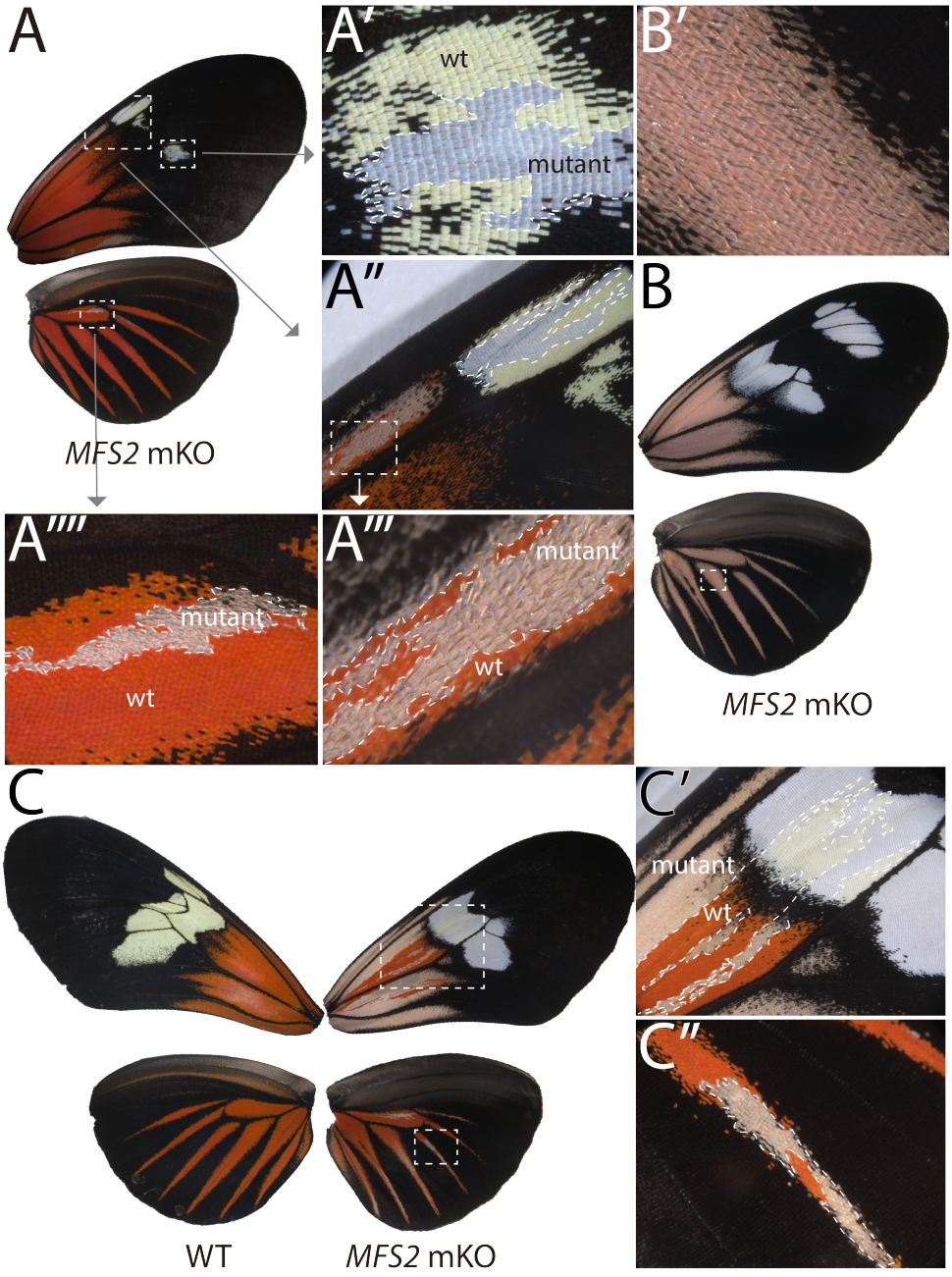

Figure 3.3 Mutant phenotypes following deletion of *MFS2*. (A) An *MFS2*-injected adult displaying mosaic KO (mKO) events, with red-to-pink mutant clones displaying a pointed scale morphology (A'',A''',A'''') and extending into the forewing "band", shown by the yellow-to-white color transition (A'). (B) An *MFS2*-injected adult displaying a full red-to-pink color transition throughout the fore- and hindwings. Note the pointed scale morphology seen in the HW ray (B'). (C) Comparison of a WT adult (left) to an *MFS2*-injected mKO adult. Insets show detail of the mutant clones extending from the red "dennis" to the yellow "band" (C') and clonal effects in the HW rays (C'').

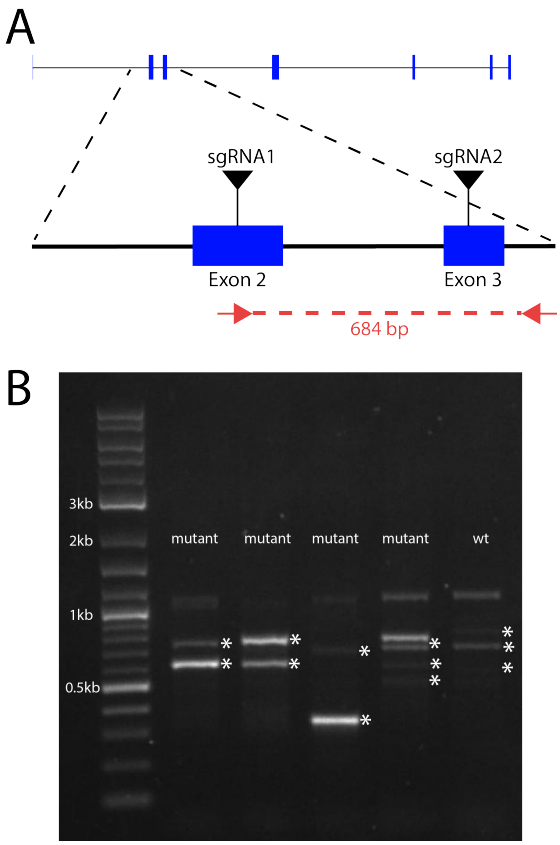

Figure 3.4 Experimental plan and genotyping results for *MFS2*. (A) A diagram showing the location of sgRNA binding sites in the *MFS2* coding region. Exons are represented by blue boxes and genotyping primers are represented by red arrows. The PCR amplified region for genotyping is highlighted in red. (B) The PCR products of 4 *MFS2*-injected adults (labeled "mutant") following amplification of the CRISPR-targeted region. Asterisks indicate bands that were gel extracted, cloned, and sequenced. All sequences obtained were unintended misamplifications of various genomic regions.

DISCUSSION

Based on previously published research supporting the role of *cn* in insect eye pigmentation, I expected *cn* mutants to exhibit a loss of red pigmentation in the wing patterns and possibly show red pigmentation in the eyes (Wald, 1946); however, after 57 egg injections yielding 16 adults, no mutant phenotypes were observed. Furthermore, genotyping confirmation of the sgRNA activity was inconclusive. These results bring ambiguity to any potential conclusions that can be made regarding the role of *cn* in *Heliconius* butterfly wing pigment synthesis. The lack of mutant phenotypes can be a result of a few different possibilities. First, the 16 butterflies that survived the injections might not have been mutant for *cn.* The sgRNAs used to target the *cn* coding region could bind poorly to their target, resulting in low efficiency or complete failure of genome editing events. To verify this hypothesis, in vitro Cas9 digestions should be completed with the RNP complex used in this study to verify its capacity for inducing double-

stranded cuts in the *H.e.lat* genome. If mutations had actually been created with the sgRNAs, a second possible explanation for the lack of mutant phenotypes could be that the mutations might not have disrupted the function of *cn*. In an attempt to target predicted functional domains, I designed the sgRNAs to target the 3' region of the *cn* gene—thus avoiding the potential for an early frameshift event that could result in complete loss of function of the protein (Fig. 3.2A). This decision may have resulted in a partially- or even fully functioning gene product and therefore may explain the lack of mutant phenotypes. For future attempts at *cn* knockouts, I suggest targeting earlier exons in the 5' region of the coding sequence. A final explanation for the lack of mutant phenotypes is that, despite evidence to the contrary, *cn* simply is not part of the wing ommochrome synthesis pathway. Although unlikely, this option cannot be ruled out from the experiments presented here in this study.

The major facilitator superfamily (MFS) of transporters is present in many species, ranging from bacteria to eukaryotes, and has been shown to be associated with ommochrome biosynthesis in egg, eye, and larval development in the silkworm *Bombyx mori* (Ito et al., 2012; Osanai-futahashi et al., 2012; Zhao et al., 2012). Recently, several MFS transporters were associated with both red and black color patterns in the wings of the butterfly *V.cardui,* suggesting their potential involvement in both melanin and ommochrome synthesis pathways (Zhang et al., 2017). Although previous research suggests MFS as having multiple roles in various pigment types across insects, the functional role of these genes in butterfly wing pigment synthesis had yet to be validated. The results presented in this study show strong support for at least one member of the MFS family, *MFS2*, as having a role in ommochrome pigmentation in the *Heliconius* ommochrome pathway. *MFS2* mutants displayed strong clonal phenotypes with significant loss of red pigmentation throughout the forewings and hindwings of the adult butterfly. Interestingly, many of these mutants also displayed a morphological change in scale structure—with depigmented scales showing a pointed shape when compared to WT

scales—suggesting *MFS2* as having a role not only in pigmentation but also potentially in scale morphology. In addition to the red pigment abnormalities observed, many mutant adults displayed a yellow-to-white color transition in the "band" element of the adult forewings. Yellow pigment in *Heliconius* butterflies is thought to be a result of the uptake of 3-OHK, an intermediate in the ommochrome synthesis pathway, into scale cells during late pupal development (Fig. 2.1) (Reed et al., 2008). Due to the combined effect of *MFS2* mutations on both red and yellow pigmentation, these results suggest the MFS2 transporter as having a potential role in the transportation of 3-OHK into scale cells. Due to the lack of genotype confirmation, however, further experiments need to be completed in order to confirm the mutations observed in this study are truly a result of deleting the intended region of *MFS2* before the results presented in this study can be considered valid.

TABLES

Table 3.1: sgRNAs and primers used in this study

Target	sgRNA Name	sgRNA sequence	Genotyping Primers
cn	helat_cn_1	GGGTATGAACGCCGGCTTCGAGG	TGTCGCCCTTATCATGTGCA
	helat_cn_2	CCCGTCCATCATATCGTATTCGG	TTCTGCCATTGACGATTCTTTACG
MFS2	helat_MFS2_1	CCTGCGATGGTGACGTCAGCAGT	TATGGCCGACGAGCTGTCA
	helat_MFS2_2	GTATCTCGGTGAGATGCACACGG	TTGCCGCGTTGCCAAAGT

Table 3.2: Injection summary

Element	sgRNAs used (total)	[sgRNA(each): Cas9] ng/µl	Injected eggs	Hatched eggs	Hatch ratio	Total adults	Mutants
cinnabar	helat_cn_1 and helat_cn_2 (2)	125:250	57	25	44%	16	0
MFS2	helat_MFS2_1 ad helat_MFS2_2 (2)	166:333	231	71	31%	37	8

REFERENCES

Ferguson, L. C., & Jiggins, C. D. (2009). Shared and divergent expression domains on mimeticHeliconiuswings. Evolution & Development, 11(5), 498-512. doi:10.1111/j.1525-142x.2009.00358.x

Gilbert, L. E., H.S. Forest, T.D. Schultz, and D. J. Harvey. (1988). Correlations of ultrastructure and pigmentation suggest how genes control development of wing scales in *Heliconius* butterflies. Journal for Research on the Lepidoptera, 26, 141-160.

Ito, K., K. Kidokoro, S. Katsuma, T. Shimada, K. Yamamoto et al. (2012) Positional cloning of a gene responsible for the cts mu- tation of the silkworm, Bombyx mori. Genome 55: 493–504.

Jiggins, C. D. (2018). Eecology and Evolution of Heliconius Butterflies. S.l.: Oxford Univ Press.

Linzen, B. (1974). The Tryptophan → Omrnochrome Pathway in Insects. Advances in Insect Physiology, 117-246. doi:10.1016/s0065-2806(08)60130-7

Nijhout, H. F. (1991). The development and evolution of butterfly wing patterns. Washington: Smithsonian Inst. Press.

Osanai-Futahashi, M., K.-I. Tatematsu, K. Yamamoto, J. Narukawa, K. Uchino et al. (2012) Identification of the Bombyx red egg gene reveals involvement of a novel transporter family gene in late steps of the insect ommochrome biosynthesis pathway. J. Biol. Chem. 287: 17706–17714.

Reed, R. D., Mcmillan, W. O., & Nagy, L. M. (2008). Gene expression underlying adaptive variation in Heliconius wing patterns: Non-modular regulation of overlapping cinnabar and vermilion prepatterns. Proceedings of the Royal Society B: Biological Sciences, 275(1630), 37-46. doi:10.1098/rspb.2007.1115

Wald, G. (1946). Fractionation Of The Eye Pigments Of Drosophila Melanogaster. The Journal of General Physiology, 30(1), 41-46. doi:10.1085/jgp.30.1.41

Zhang, L., Martin, A., Perry, M. W., Karin R. L. Van Der Burg, Matsuoka, Y., Monteiro, A., & Reed, R. D. (2017). Genetic Basis of Melanin Pigmentation in Butterfly Wings. Genetics, 205(4), 1537-1550. doi:10.1534/genetics.116.196451

Zhang L., Reed R.D. (2017) A Practical Guide to CRISPR/Cas9 Genome Editing in Lepidoptera. In: Sekimura T., Nijhout H. (eds) Diversity and Evolution of Butterfly Wing Patterns. Springer, Singapore

Zhao, Y., H. Zhang, Z. Li, J. Duan, J. Jiang et al. (2012) A major facilitator superfamily protein participates in the reddish brown pigmentation in Bombyx mori. J. Insect Physiol. 58: 1397–1405.